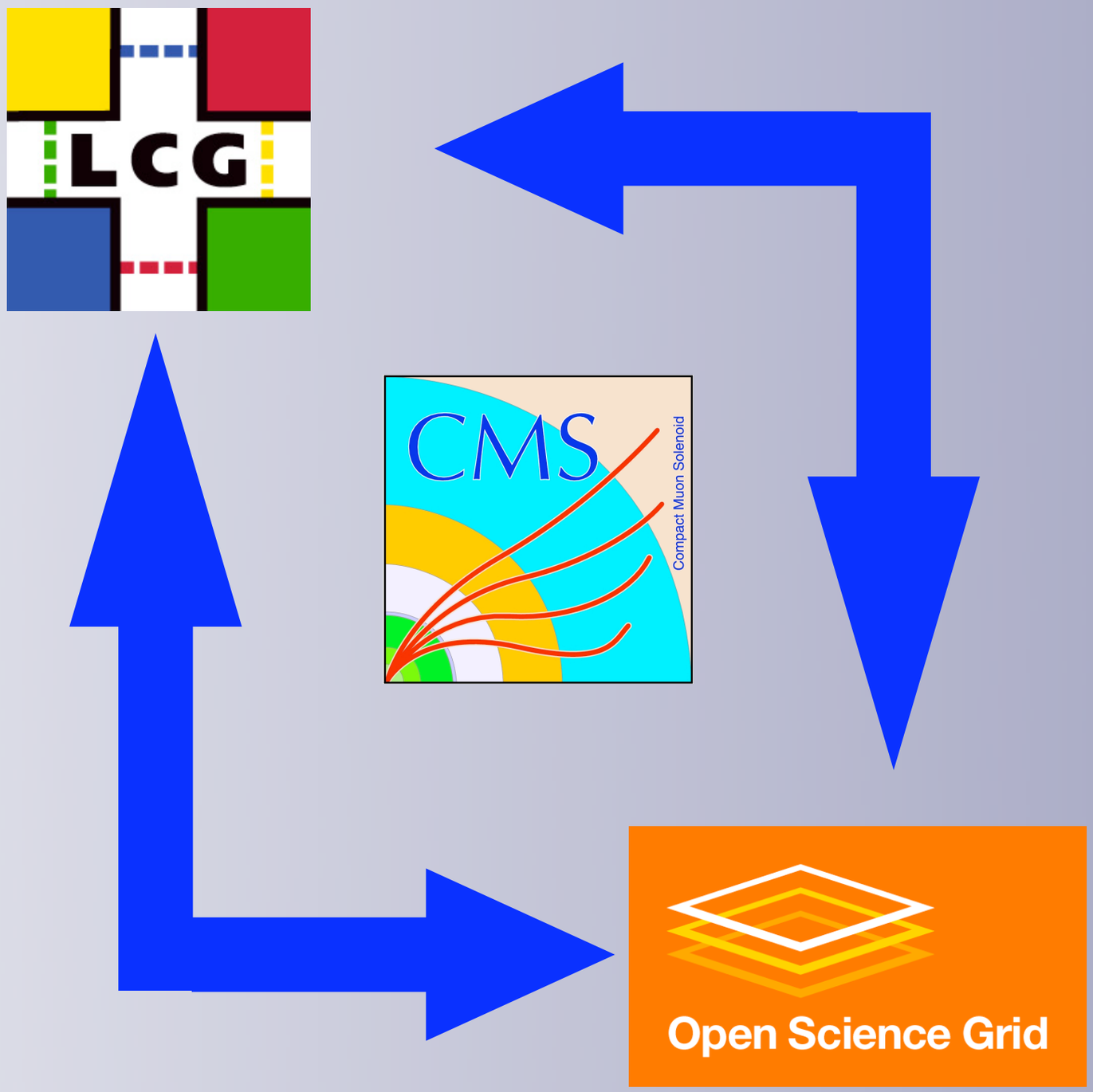
CMS Workflow Activities on OSG

CMS Activity Coordination Meeting
11 / 22 / 05

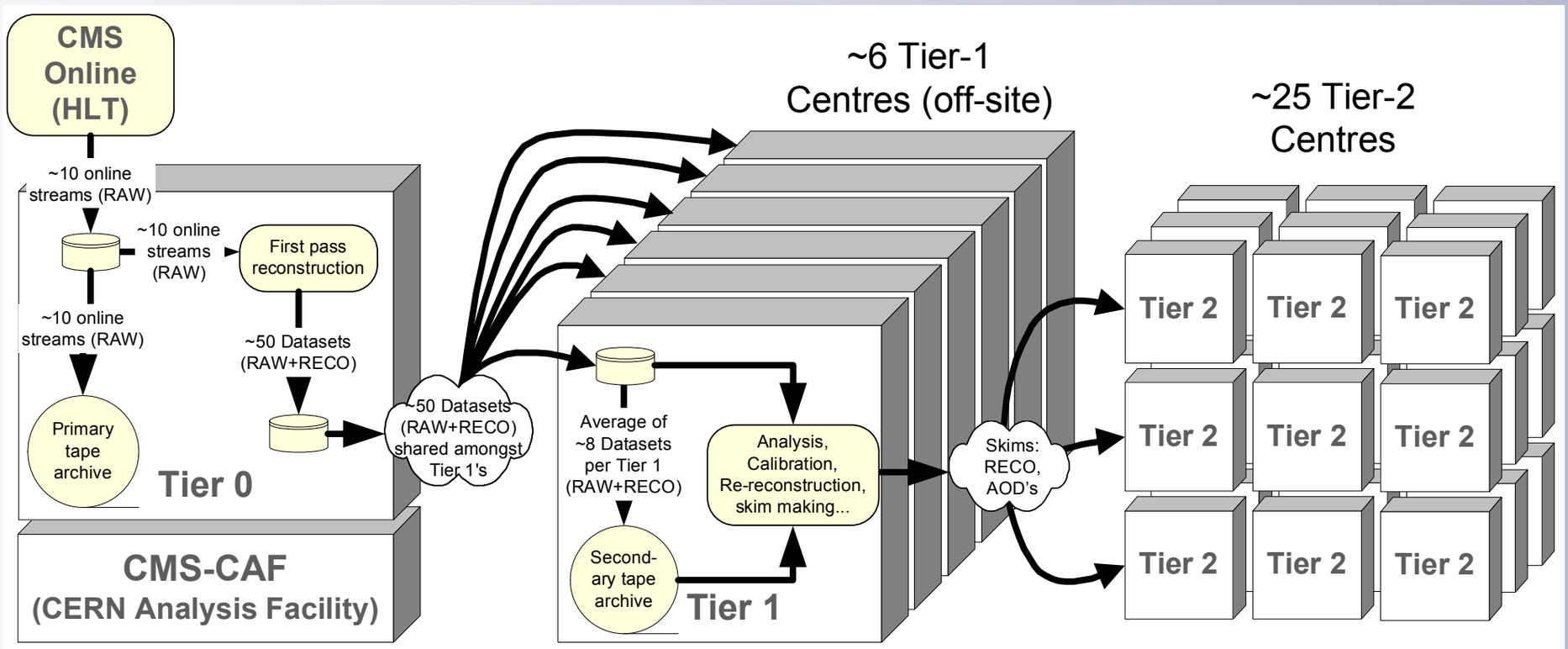
Oliver Gutsche
USCMS / Fermilab

Outline

- **CRAB** - **C**MS **R**emote **A**nalysis **B**uilder
  - Introduction
  - Functionality
  - Dataset discovery
  - Statistics
- LCG / OSG
  - Overview US OSG T2 centers
- **OSG additions to CRAB**
- Service Challenge 3
  - Statistics
- Operational Experience
  - current Event Data Model
  - Dataset Transfer and Publication



CMS Data Model



- CMS Data Analysis (and MC production)
  - worldwide distributed
    - Tier-Structure, each hosting datasets and providing analysis facilities
    - T1 and T2 represent significant computing resources
- All need to be accessible via GRID interfaces for all CMS users

		Running Year 2007	2008	2009	2010	
Conditions		Pilot	2E33+HI	2E33+HI	E34+HI	
Tier-0	CPU	2.3	4.6	6.9	11.5	MSi2k
	Disk	0.1	0.4	0.4	0.6	PB
	Tape	1.1	4.9	9	12	PB
	WAN	3	5	8	12	Gb/s
A Tier-1	CPU	1.3	2.5	3.5	6.8	MSi2k
	Disk	0.3	1.2	1.7	2.6	PB
	Tape	0.6	2.8	4.9	7.0	PB
	WAN	3.6	7.2	10.7	16.1	Gb/s
Sum Tier-1	CPU	7.6	15.2	20.7	40.7	MSi2k
	Disk	2.1	7.0	10.5	15.7	PB
	Tape	3.8	16.7	29.5	42.3	PB
A Tier-2	CPU	0.4	0.9	1.4	2.3	MSi2k
	Disk	0.1	0.2	0.4	0.7	PB
	WAN	0.3	0.6	0.8	1.3	Gb/s
Sum Tier-2	CPU	9.6	19.3	32.3	51.6	MSi2k
	Disk	1.5	4.9	9.8	14.7	PB
CMS CERN Analysis Facility (CMS-CAF)	CPU	2.4	4.8	7.3	12.9	MSi2k
	Disk	0.5	1.5	2.5	3.7	PB
	Tape	0.4	1.9	3.3	4.8	PB
	WAN	0.3	5.7	8.5	12.7	Gb/s
Total	CPU	21.9	43.8	67.2	116.6	MSi2k
	Disk	4.1	13.8	23.2	34.7	PB
	Tape	5.4	23.4	41.5	59.5	PB

CRAB

- Access to dataset for distributed analysis
  - **CRAB** - **C**MS **R**emote **A**nalysis **B**uilder
- Provides CMS users with
  - framework to run their analysis on datasets hosted by CMS T1 and T2 centers
    - No detailed knowledge about GRID infrastructures necessary
- Uses GRID infrastructure
  - Authentication by GRID certificates and virtual organizations (VO's)
  - Job interaction (submission, status request, output retrieval) using GRID middleware

CRAB - a short introduction

- CRAB splits User interaction into steps:
  - Creation of Jobs
  - Submission of Jobs
  - Status check of Jobs
  - Retrieval of Job output
- CRAB takes care of User code:
  - Packing of User executable and libraries
  - Shipping of User code to worker node (WN) for execution
  - Preparation of Software environment on WN and execution

Creation: data discovery

User:
request to analyze
dataset with user code

1. resolve requested dataset into identifier

2. inquire which centers publish requested dataset

3. contact centers and inquire about dataset locally

local catalog

local catalog

local catalog

T1

T1

T2

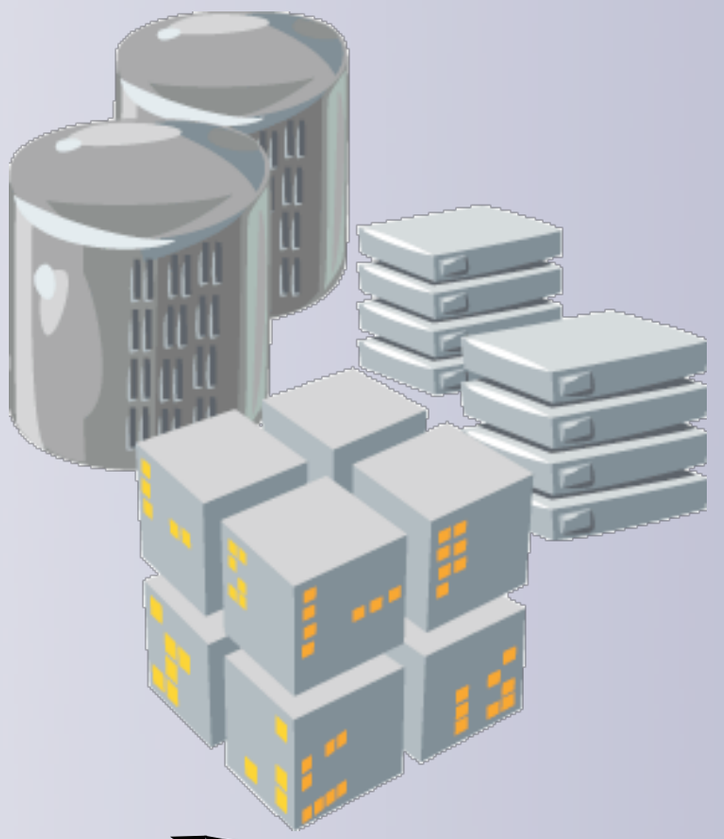
Jobs are created locally

- on the User's submission computer
- each job is able to run on all centers from the request list

Submission, Status inquiry and Output retrieval

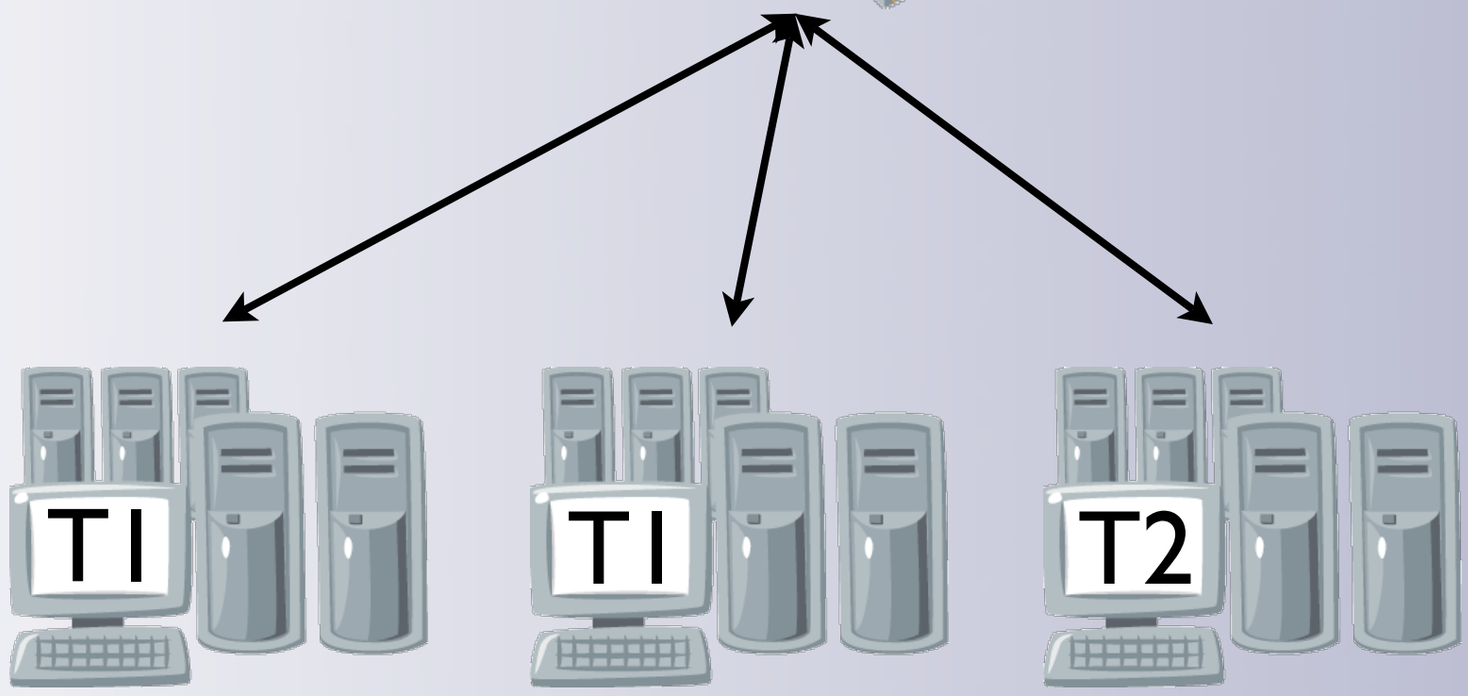
User's submitter

- providing created jobs to RB (input sandbox)
- checking status of jobs
- retrieving output (output sandbox)



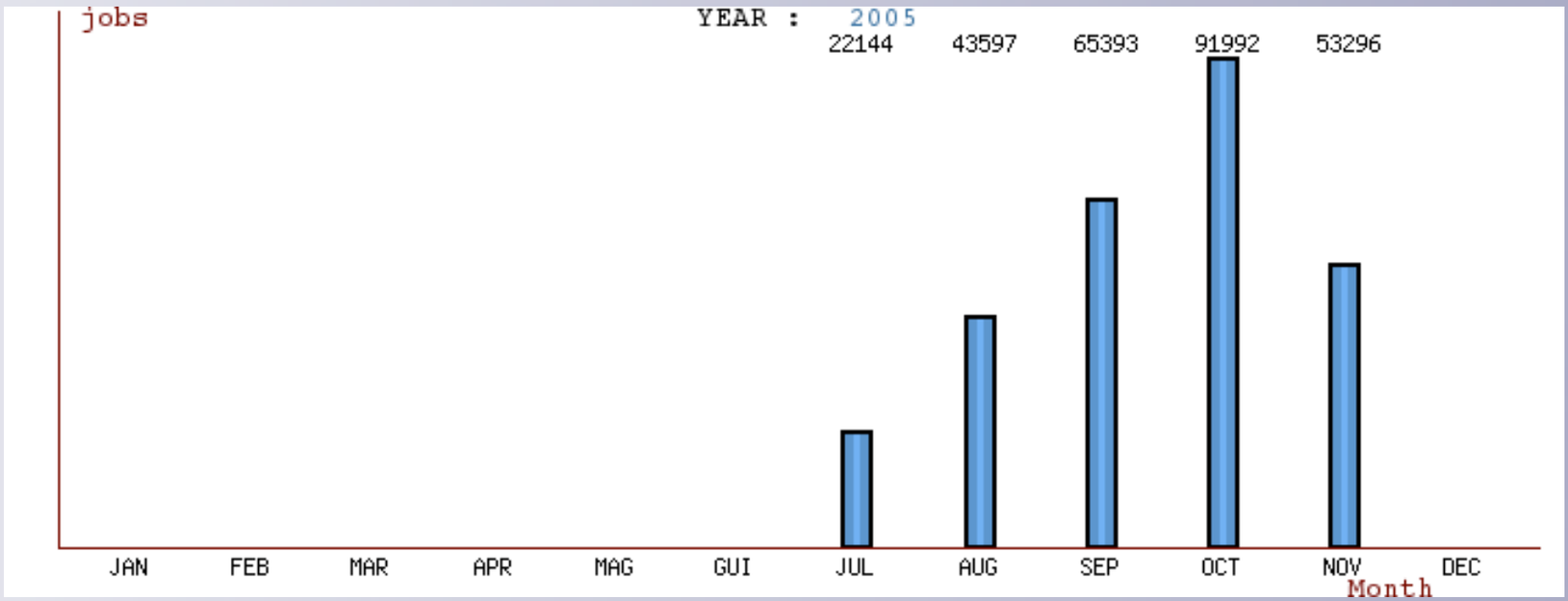
Resource Broker (RB)

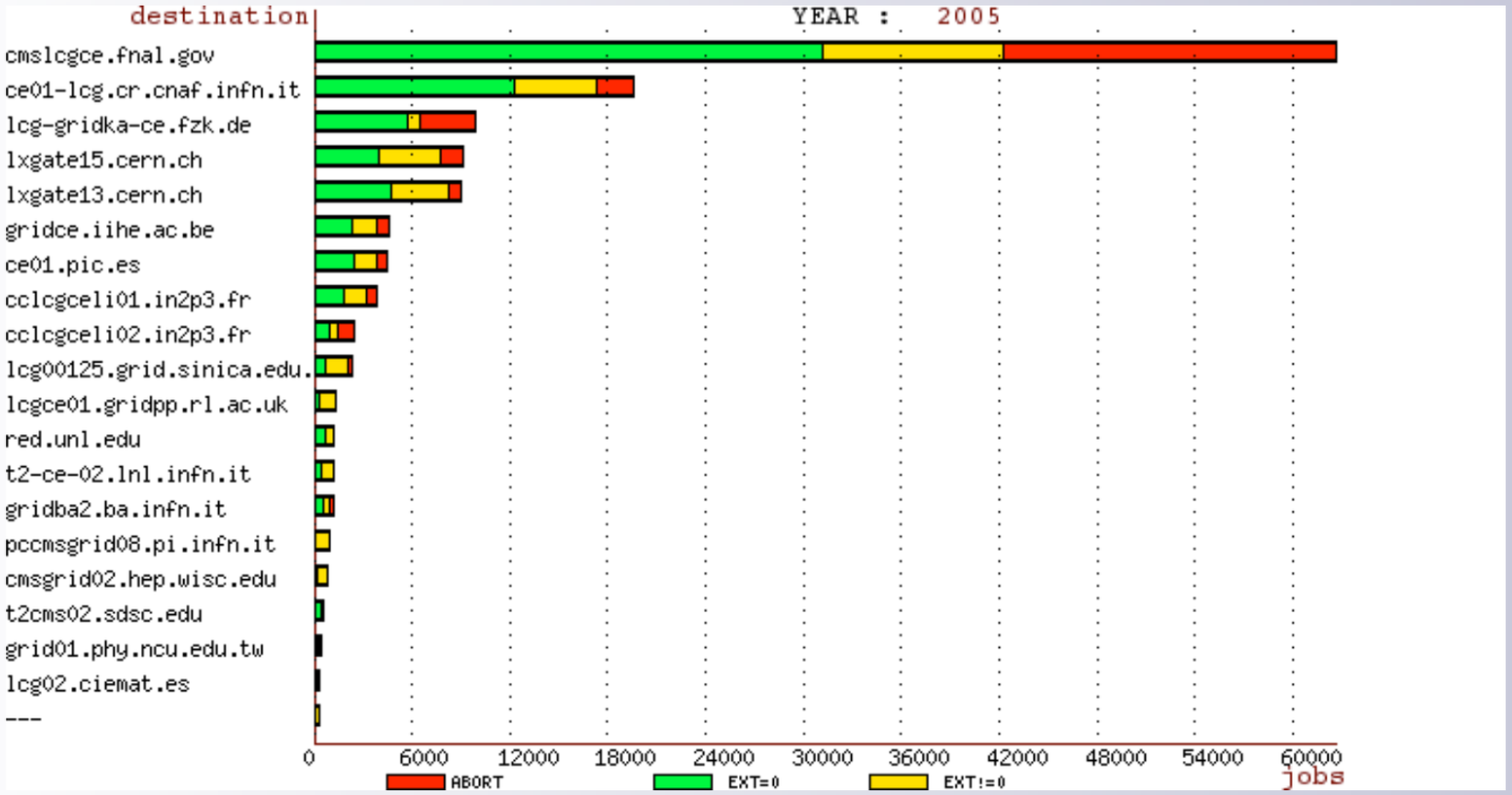
- brokers job between requested centers
- provides input and output sandbox for file handling



Status

- increasing usage within the CMS analysis community

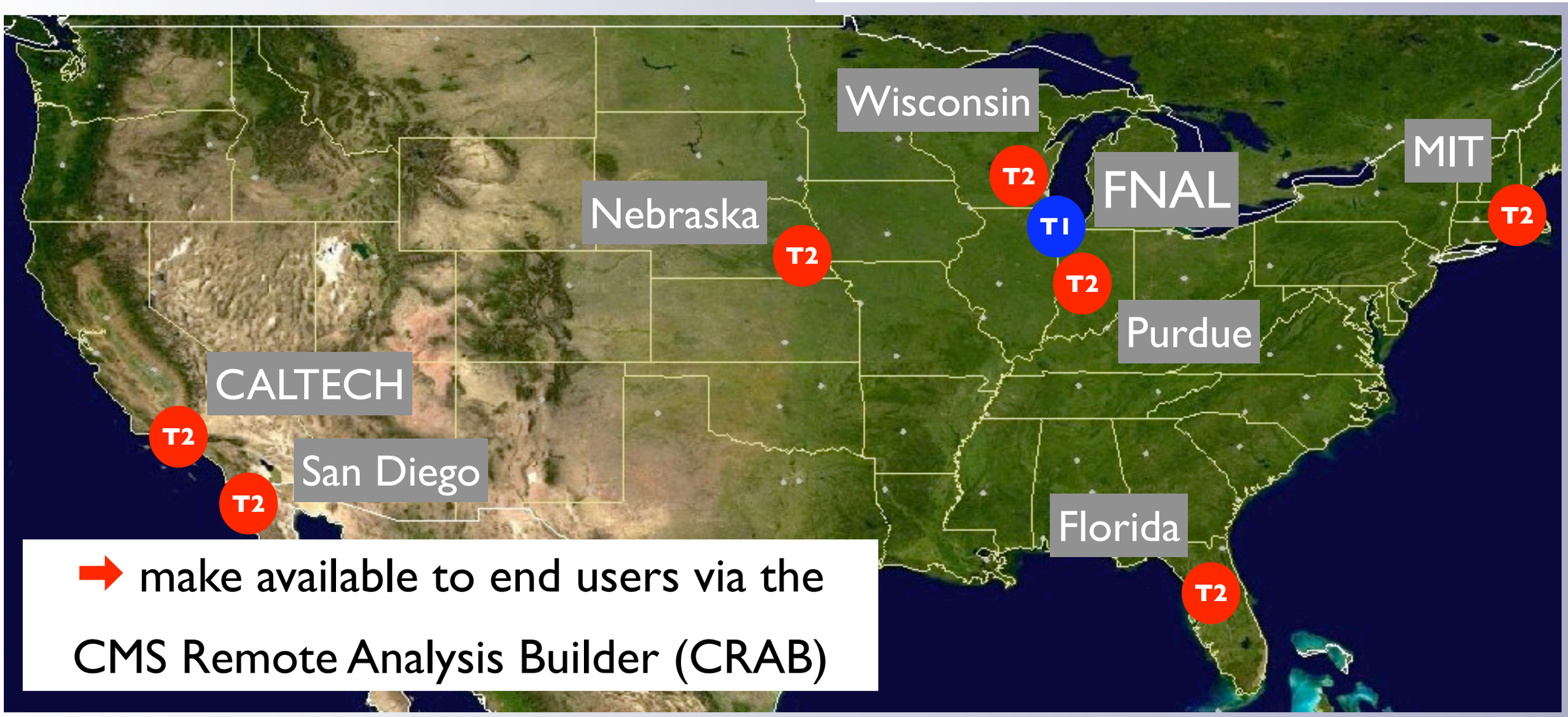




US contribution to CMS Tier structure

- U.S. contribution to CMS tier structure
  - T1 at FNAL providing LHC Computing Grid (LCG) and OpenScience Grid (OSG) interfaces
  - 7 attached T2 sites using OSG infrastructure

Site	Processors	Disk (TB)
Caltech	153	40
Florida	240+	73
MIT	(coming soon)	(coming soon)
Nebraska	256	19
Purdue	228	~25
San Diego	228	44.5
Wisconsin	400	50



➡ make available to end users via the CMS Remote Analysis Builder (CRAB)

CRAB and OSG

- **CRAB** based on LCG / EDG middleware using more higher level tools
  - access via EDG tools like **edg-job-submit**
  - utilization of Resource Broker (RB)
    - load balancing
    - sandbox for user file input and output to the remote analysis application
- **OSG** based on VDT suite providing GLOBUS toolkit using more lower level tools
  - access via GLOBUS tools like **globus-job-submit**
  - no RB
  - missing sandbox functionality

➡ CRAB cannot be used directly

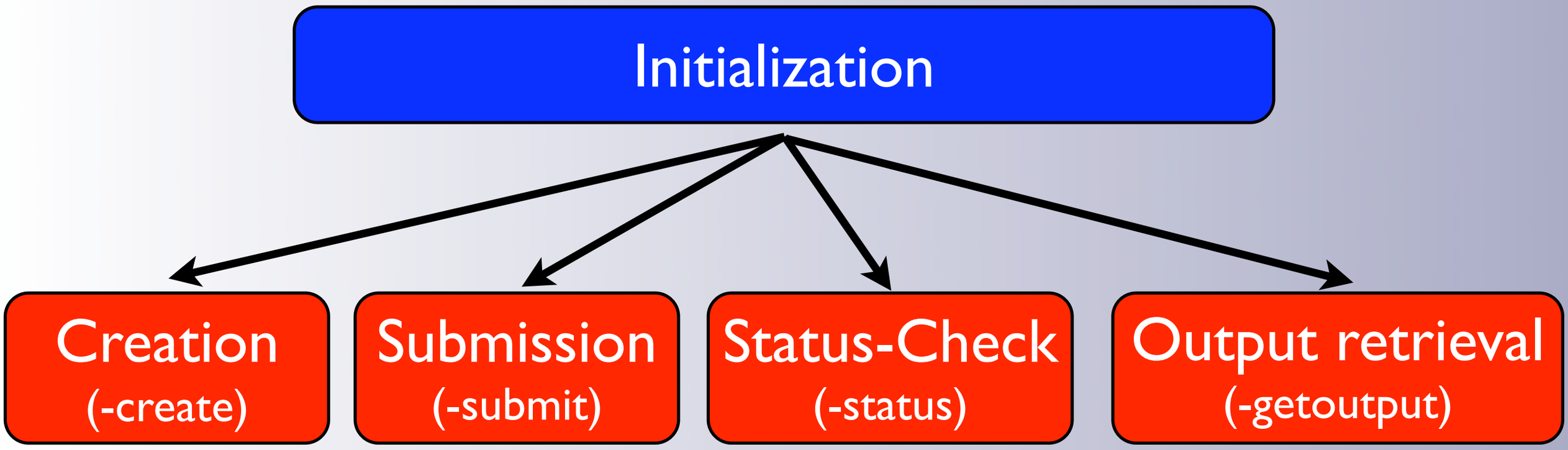
➡ **Add** functionality to CRAB to be able to also submit to OSG sites

CONDOR-G

- First approach: CONDOR-G provides:
  - GRID submission functionalities using GLOBUS toolkit
  - access to OSG sites independent of used local batch system
  - sandbox for insertion and retrieval of files
- Requirements:
  - OSG T2 site:
    - **none**
  - Submitter:
    - local CONDOR installation with activated CONDOR-G

Implementation

- enable CRAB to **identify OSG sites** for requested dataset
- first approach integrating concept of OSG submission transparent into CRAB:
  - ➡ OSG mode with hardcoded information
    - OSG T2 sites PubDB URL's
    - batch system of OSG T2 sites for jobmanager identification (EDG: RB, BDII)
    - path to CMS software installation
- CRAB decomposition:



Initialization

First (follow up uses conf. file)

parse options
(file and command line)

create directories and store configuration

create job type

for requested dataset/owner
- find collection id's
- find PubDB's publishing data

OSG mode

in "check PubDB list"
- compare to hardcoded OSG list
- keep only OSG sites
- in the following, take the first

Creation

Creation

write JDL's

OSG mode

- information content the same, structure of CONDOR-G JDL completely different
- take CE of first selected OSG PubDB
- use corresponding hardcoded jobmanager

write job execution script

OSG mode

- use _CONDOR_SCRATCH_DIR where appropriate
- source setup script from hardcoded CMS software path
- use first selected OSG site for init script (catalog download) and orcarc site dependent fragment

Submission & Status & Getoutput

Submission

use edg-job-submit

OSG mode
- use condor_submit

Status

use edg tools

OSG mode
- use condor_q

Submission

use edg-job-getoutput

OSG mode
- CONDOR-G does not need a trigger for output retrieval

Service Challenge 3

- tests dataset transfer from T0 to T1 and subsequent T2's
  - validate datasets at T1 and T2 using CRAB
  - OSG modified version is used for US T2's
- 2nd phase started Nov. 14, statistics so far:

Name	CALTECH	PURDUE	SDSC	UFL	UNL	WISC	All OSG T2 Sites
Jobs successful	0	0	487	5	0	172	664
Job with non-zero status	20	0	10	0	0	276	306
All Completed jobs	20	0	502	9	0	448	979

SC3 Experience: EDM

- goal of SC3 CRAB job efforts:
  - validate transferred dataset using the old Event Data Model (ORCA)
- Experiences with old EDM:
  - unreliable execution of jobs on Digi level
    - frequent crashes
    - program termination by underlying framework
  - impossible execution of jobs on DST level
    - no successful jobs at all at OSG T2's
    - excluded from the Service Challenge

SC3 Experience: Dataset Transfer and Publication

- Datasets are distributed from T0 to T1 and T2 using PhEDEx (Physics Experiment Data Export)
  - Transfer agents manage movement of files between sites
- Prompt Publication after transfer is handled by CMSGLIDE
- Experience
  - heavily dependent on performance of mass storage system (Castor 2 at CERN, dCache at FNAL and OSG T2's)
  - instabilities in Transfer agents:
    - need a lot of attention by the site admins to achieve good transfer rates
  - Complicated Publication procedure:
    - after arriving at site, METADATA of dataset has to be "attached" to local METADATA catalog
      - fails due to instabilities of EDM (ORCA)
      - problems when used EDM versions at generation and attach do not match
    - local site configuration problems

Summary & Outlook

- First OSG implementation in CRAB
  - submit analysis jobs to OSG T2's
  - participate in Service Challenge 3
- Plans:
  - OSG features of CRAB are currently only available to experts
  - new version of CRAB (1.0) released
    - OSG features for all CMS users are planned to be implemented here
    - plan to use the RB rather than direct submission